Deep learning models for brain imaging
model depth enhances discovery power

Sergey M. Plis

The Mind Research Network For Neurodiagnostic Discovery
Lovelace
The University of New Mexico

joint work with Devon Hjelm, Ruslan Salakhutdinov, Jeremy Bockholt, Jeffrey Long, Hans Johnson, Jane Paulsen Jessica Turner and Vince D. Calhoun

Why deep learning

- Deep learning was created for automatic representation[1]:
  - roots in artificial neural nets (representation machines)
  - can work in generative mode
  - learn from the data (feature selection not required)
- Deep learning is beating records in pattern recognition:
  - natural image classification[2]
  - street sign recognition (superhuman performance)
  - for more see `http://tinyurl.com/os3xnpb`
- Widely successful in commercial pattern recognition
  - Google's Deep Brain division
  - Facebook's deep learning department
  - Microsoft, Amazon, Yahoo, Baidu, etc.
- Extensive press coverage: New York Times (twice!), Wired...
- Not yet widely used in natural sciences
- Until recently (are we the first?), close to zero use in neuroimaging

[1] Hinton, G. *et al.* *Science* **313,** 504–507 (2006).

[2] Krizhevsky, A. *et al.* in *Neural Information Processing Systems* (2012).

Why deep learning

- Deep learning was created for automatic representation[1]:
  - roots in artificial neural nets (representation machines)
  - can work in generative mode
  - learn from the data (feature selection not required)
- Deep learning is beating records in pattern recognition:
  - natural image classification[2]
  - street sign recognition (superhuman performance)
  - for more see `http://tinyurl.com/os3xnpb`
- Widely successful in commercial pattern recognition
  - Google's Deep Brain division
  - Facebook's deep learning department
  - Microsoft, Amazon, Yahoo, Baidu, etc.
- Extensive press coverage: New York Times (twice!), Wired...
- Not yet widely used in natural sciences
- Until recently (are we the first?), close to zero use in neuroimaging

[1] Hinton, G. *et al.* *Science* **313,** 504–507 (2006).

[2] Krizhevsky, A. *et al.* in *Neural Information Processing Systems* (2012).

Why deep learning

- Deep learning was created for automatic representation[1]:
  - roots in artificial neural nets (representation machines)
  - can work in generative mode
  - learn from the data (feature selection not required)
- Deep learning is beating records in pattern recognition:
  - natural image classification[2]
  - street sign recognition (superhuman performance)
  - for more see `http://tinyurl.com/os3xnpb`
- Widely successful in commercial pattern recognition
  - Google's Deep Brain division
  - Facebook's deep learning department
  - Microsoft, Amazon, Yahoo, Baidu, etc.
- Extensive press coverage: New York Times (twice!), Wired...
- Not yet widely used in natural sciences
- Until recently (are we the first?), close to zero use in neuroimaging

[1] Hinton, G. *et al.* *Science* **313,** 504–507 (2006).

[2] Krizhevsky, A. *et al.* in *Neural Information Processing Systems* (2012).

Why deep learning

- Deep learning was created for automatic representation[1]:
  - roots in artificial neural nets (representation machines)
  - can work in generative mode
  - learn from the data (feature selection not required)
- Deep learning is beating records in pattern recognition:
  - natural image classification[2]
  - street sign recognition (superhuman performance)
  - for more see `http://tinyurl.com/os3xnpb`
- Widely successful in commercial pattern recognition
  - Google's Deep Brain division
  - Facebook's deep learning department
  - Microsoft, Amazon, Yahoo, Baidu, etc.
- Extensive press coverage: New York Times (twice!), Wired...
- Not yet widely used in natural sciences
- Until recently (are we the first?), close to zero use in neuroimaging

[1] Hinton, G. *et al.* *Science* **313,** 504–507 (2006).

[2] Krizhevsky, A. *et al.* in *Neural Information Processing Systems* (2012).

Why deep learning

- Deep learning was created for automatic representation[1]:
  - roots in artificial neural nets (representation machines)
  - can work in generative mode
  - learn from the data (feature selection not required)
- Deep learning is beating records in pattern recognition:
  - natural image classification[2]
  - street sign recognition (superhuman performance)
  - for more see `http://tinyurl.com/os3xnpb`
- Widely successful in commercial pattern recognition
  - Google's Deep Brain division
  - Facebook's deep learning department
  - Microsoft, Amazon, Yahoo, Baidu, etc.
- Extensive press coverage: New York Times (twice!), Wired...
- Not yet widely used in natural sciences
- Until recently (are we the first?), close to zero use in neuroimaging

[1] Hinton, G. *et al.* *Science* **313,** 504–507 (2006).

[2] Krizhevsky, A. *et al.* in *Neural Information Processing Systems* (2012).

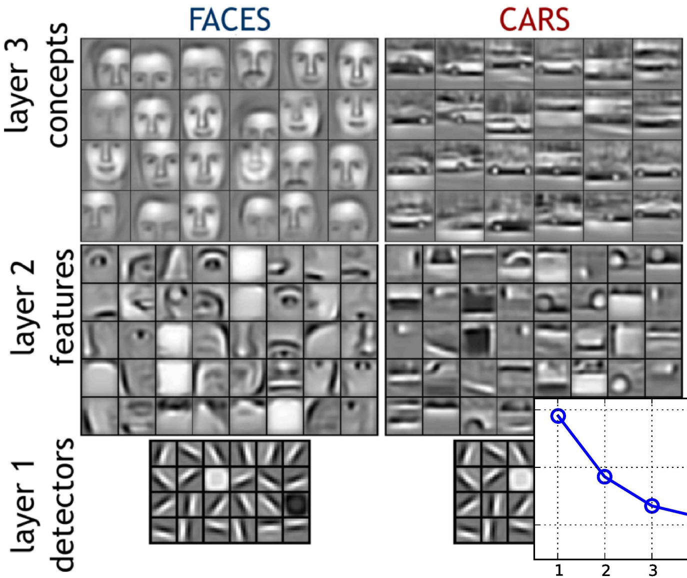
Why deep learning

- Deep learning was created for automatic representation[1]:
  - roots in artificial neural nets (representation machines)
  - can work in generative mode
  - learn from the data (feature selection not required)
- Deep learning is beating records in pattern recognition:
  - natural image classification[2]
  - street sign recognition (superhuman performance)
  - for more see `http://tinyurl.com/os3xnpb`
- Widely successful in commercial pattern recognition
  - Google's Deep Brain division
  - Facebook's deep learning department
  - Microsoft, Amazon, Yahoo, Baidu, etc.
- Extensive press coverage: New York Times (twice!), Wired...
- Not yet widely used in natural sciences
- Until recently (are we the first?), close to zero use in neuroimaging

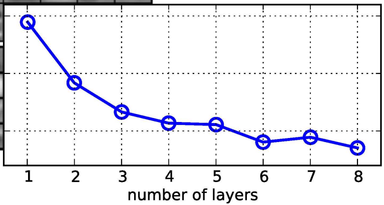
[1] Hinton, G. *et al.* *Science* **313,** 504–507 (2006).

[2] Krizhevsky, A. *et al.* in *Neural Information Processing Systems* (2012).

What is deep learning (a biased view)



- a powerful probabilistic building block
- a superior classifier
- depth increases modeling power
- an excellent representation learner



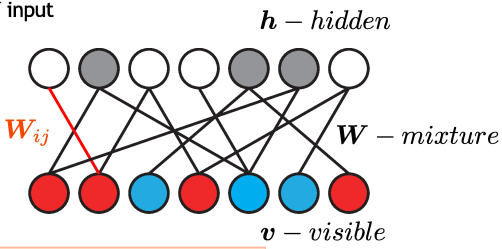
Does it hold on neuroimaging data?

Restricted Boltzmann Machine (RBM)

GOAL — estimate the joint distribution of input and hidden variables

PROXY — estimate θ that maximizes the probability that the model assigns to the data

$$\mathrm{agrmax}_\theta \prod_{v \in \boldsymbol{V}} P(v)$$



MODEL

parameters

$$\theta = \{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{W}\}$$

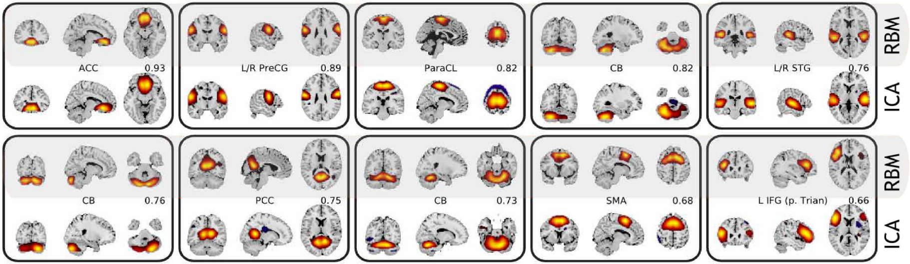
joint distribution

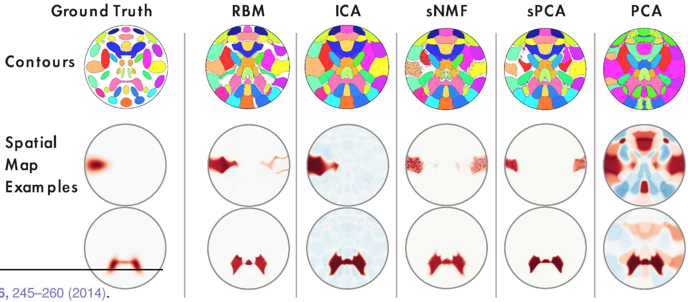
$$P(\boldsymbol{v}, \boldsymbol{h}; \theta) = \frac{1}{Z(\theta)} \exp(-E(\boldsymbol{v}, \boldsymbol{h}; \theta))$$

Gibbs energy

$$E(\boldsymbol{v}, \boldsymbol{h}; \theta) = -\sum_{i=1}^{V} \sum_{j=1}^{H} v_i W_{ij} h_j - \sum_{i=1}^{V} a_i v_i - \sum_{j=1}^{H} b_j h_j$$

RBM learns useful representations[3]

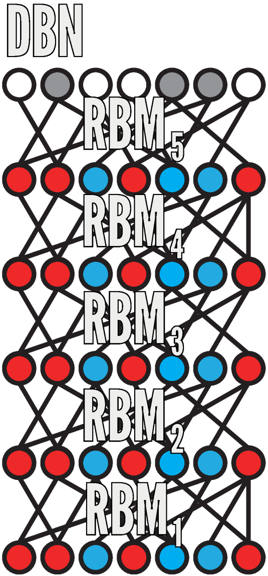




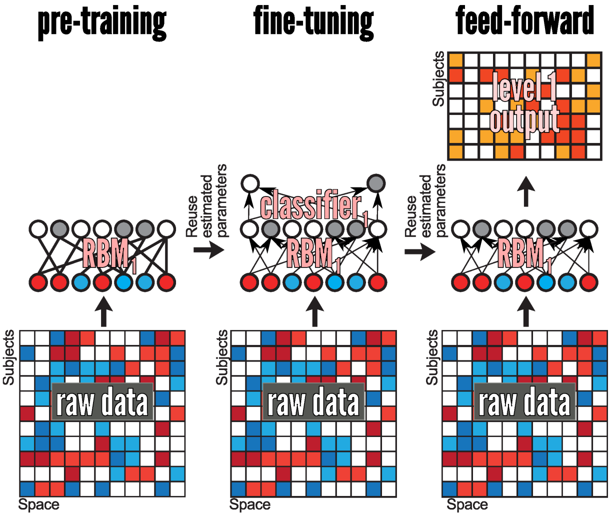
3 Hjelm, R. D. *et al.* *NeuroImage* **96,** 245–260 (2014).

Deep Belief Network

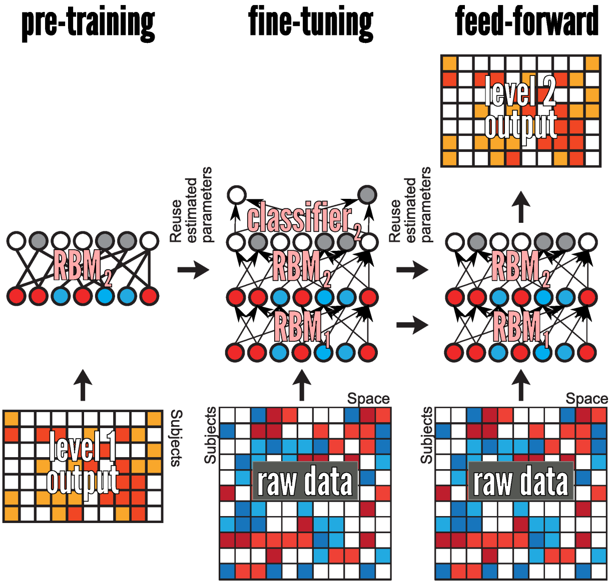
- stack RBMs
- regulate the depth
- in this work we investigate a 3 layer network
- 50-50-100 hidden units



Converting RBM training into deeper models



Converting RBM training into deeper models



Depth helps classification (numbers)

- A combined MRI dataset from schizophrenia studies at[4]:
  1. Johns Hopkins University (JHU),
  2. the Maryland Psychiatric Research Center (MPRC),
  3. the Institute of Psychiatry, London, UK (IOP), and
  4. the Western Psychiatric Institute and Clinic (WPIC).
- 198 schizophrenia patients and 191 matched healthy controls;
- 10-fold cross validation of classification performance

[4] Meda, S. A. *et al.* *Schizophrenia research* **101,** 95–105 (2008).

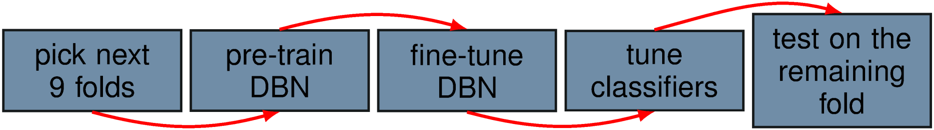
Depth helps classification (numbers)

- A combined MRI dataset from schizophrenia studies at[4]:
  1. Johns Hopkins University (JHU),
  2. the Maryland Psychiatric Research Center (MPRC),
  3. the Institute of Psychiatry, London, UK (IOP), and
  4. the Western Psychiatric Institute and Clinic (WPIC).
- 198 schizophrenia patients and 191 matched healthy controls;
- 10-fold cross validation of classification performance

[4] Meda, S. A. *et al.* *Schizophrenia research* **101,** 95–105 (2008).

Depth helps classification (numbers)

- A combined MRI dataset from schizophrenia studies at[4]:
  1. Johns Hopkins University (JHU),
  2. the Maryland Psychiatric Research Center (MPRC),
  3. the Institute of Psychiatry, London, UK (IOP), and
  4. the Western Psychiatric Institute and Clinic (WPIC).
- 198 schizophrenia patients and 191 matched healthy controls;
- 10-fold cross validation of classification performance

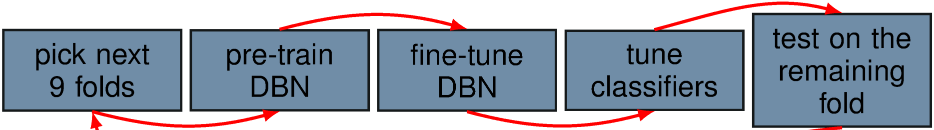


depth	raw	1	2	3
SVM F-score	0.68 ± 0.01	0.66 ± 0.09	0.62 ± 0.12	0.90 ± 0.14
LR F-score	0.63 ± 0.09	0.65 ± 0.11	0.61 ± 0.12	0.91 ± 0.14
KNN F-score	0.61 ± 0.11	0.55 ± 0.15	0.58 ± 0.16	0.90 ± 0.16

[4] Meda, S. A. *et al.* *Schizophrenia research* **101,** 95–105 (2008).

Depth helps classification (numbers)

- A combined MRI dataset from schizophrenia studies at[4]:
  1. Johns Hopkins University (JHU),
  2. the Maryland Psychiatric Research Center (MPRC),
  3. the Institute of Psychiatry, London, UK (IOP), and
  4. the Western Psychiatric Institute and Clinic (WPIC).
- 198 schizophrenia patients and 191 matched healthy controls;
- 10-fold cross validation of classification performance



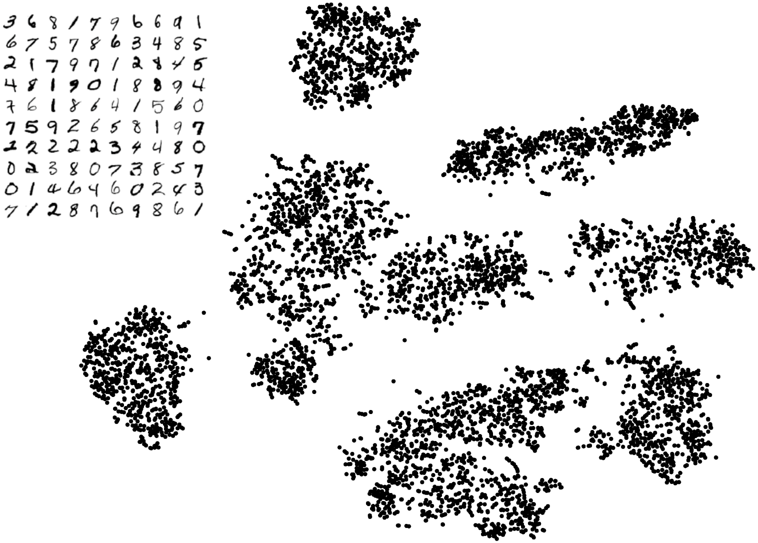
depth	raw	1	2	3
SVM F-score	0.68 ± 0.01	0.66 ± 0.09	0.62 ± 0.12	0.90 ± 0.14
LR F-score	0.63 ± 0.09	0.65 ± 0.11	0.61 ± 0.12	0.91 ± 0.14
KNN F-score	0.61 ± 0.11	0.55 ± 0.15	0.58 ± 0.16	0.90 ± 0.16

[4] Meda, S. A. *et al.* *Schizophrenia research* **101,** 95–105 (2008).

A “simple” way of making sense of the data[5]

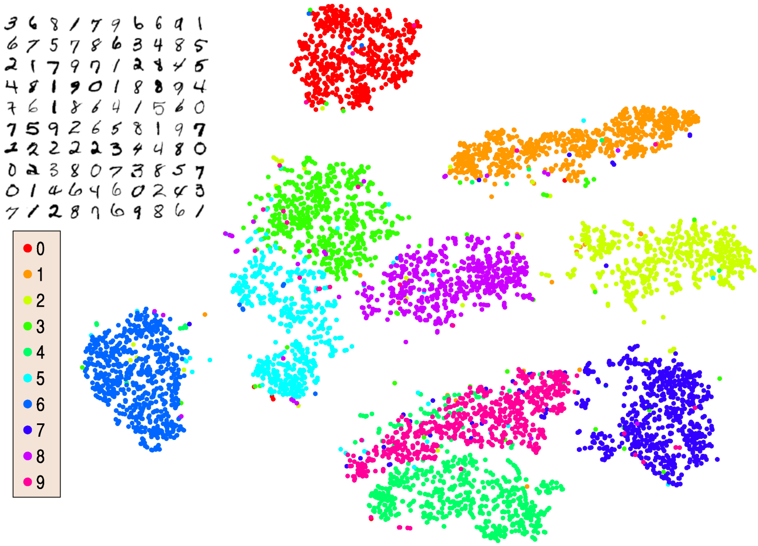
5 Van der Maaten, L. *et al.* *Journal of Machine Learning Research* **9,** 85 (2008).

A “simple” way of making sense of the data[5]



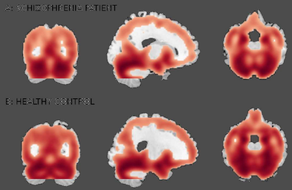
[5] Van der Maaten, L. *et al.* *Journal of Machine Learning Research* **9,** 85 (2008).

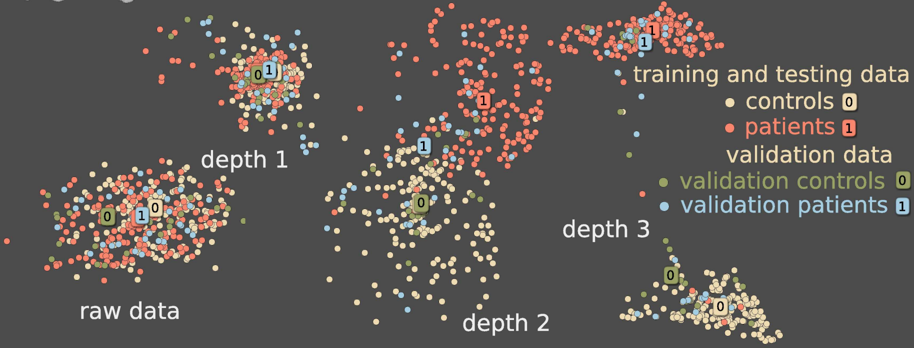
A "simple" way of making sense of the data[5]



[5] Van der Maaten, L. *et al.* *Journal of Machine Learning Research* **9,** 85 (2008).

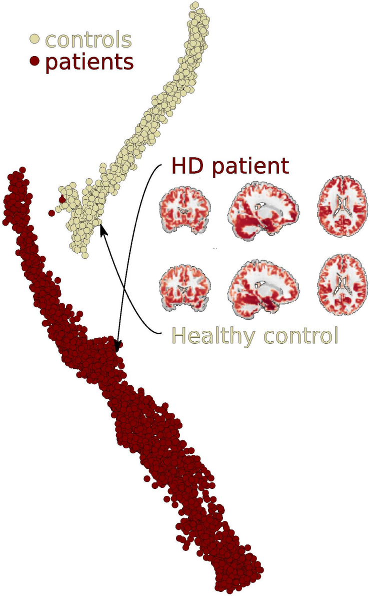
Depth helps classification (explanations)





..and given enough data depth helps discovery

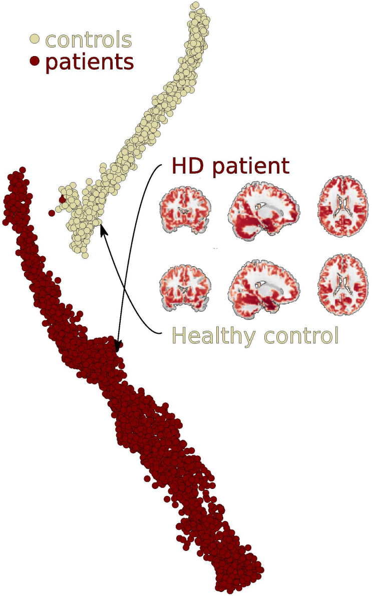
- Huntington disease data[a]
- 3500 MRI scans
- 1200 subjects
- not doing classification
- using all data and the class label (patient/control) to learn representations
- mapping 100-dimensional third level features to a 2D map
- did deepnet learn something we did not tell it?
- disease severity!



[a] www.predict-hd.net

..and given enough data depth helps discovery

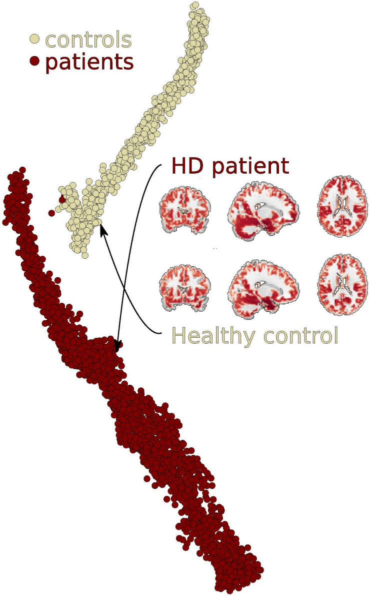
- Huntington disease data[a]
- 3500 MRI scans
- 1200 subjects
- not doing classification
- using all data and the class label (patient/control) to learn representations
- mapping 100-dimensional third level features to a 2D map
- did deepnet learn something we did not tell it?
- disease severity!



[a] www.predict-hd.net

..and given enough data depth helps discovery

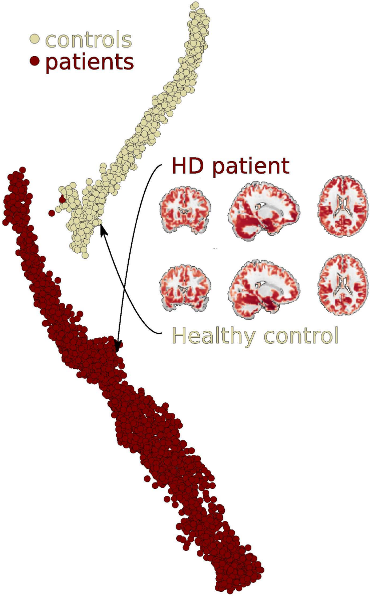
- Huntington disease data[a]
- 3500 MRI scans
- 1200 subjects
- not doing classification
- using all data and the class label (patient/control) to learn representations
- mapping 100-dimensional third level features to a 2D map
- did deepnet learn something we did not tell it?
- disease severity!



[a] www.predict-hd.net

..and given enough data depth helps discovery

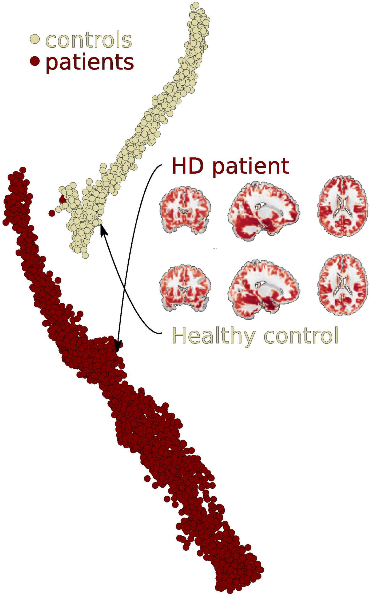
- Huntington disease data[a]
- 3500 MRI scans
- 1200 subjects
- not doing classification
- using all data and the class label (patient/control) to learn representations
- mapping 100-dimensional third level features to a 2D map
- did deepnet learn something we did not tell it?
- disease severity!



[a] www.predict-hd.net

..and given enough data depth helps discovery

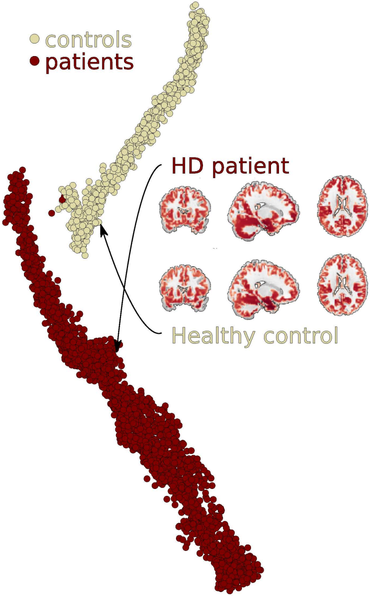
- Huntington disease data[a]
- 3500 MRI scans
- 1200 subjects
- not doing classification
- using all data and the class label (patient/control) to learn representations
- mapping 100-dimensional third level features to a 2D map
- did deepnet learn something we did not tell it?
- disease severity!



[a] www.predict-hd.net

..and given enough data depth helps discovery

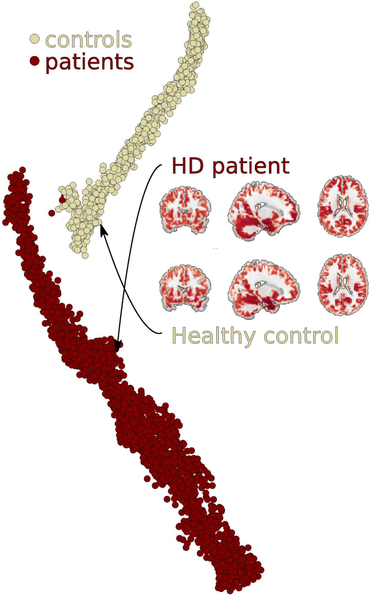
- Huntington disease data[a]
- 3500 MRI scans
- 1200 subjects
- not doing classification
- using all data and the class label (patient/control) to learn representations
- mapping 100-dimensional third level features to a 2D map
- did deepnet learn something we did not tell it?
- disease severity!



[a] www.predict-hd.net

..and given enough data depth helps discovery

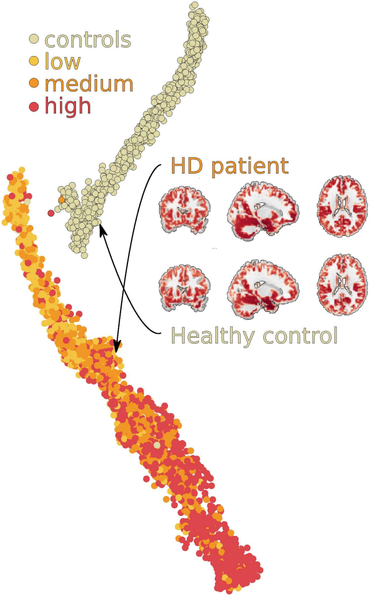
- Huntington disease data[a]
- 3500 MRI scans
- 1200 subjects
- not doing classification
- using all data and the class label (patient/control) to learn representations
- mapping 100-dimensional third level features to a 2D map
- did deepnet learn something we did not tell it?
- disease severity!



[a] www.predict-hd.net

..and given enough data depth helps discovery

- Huntington disease data[a]
- 3500 MRI scans
- 1200 subjects
- not doing classification
- using all data and the class label (patient/control) to learn representations
- mapping 100-dimensional third level features to a 2D map
- did deepnet learn something we did not tell it?
- disease severity!



[a] www.predict-hd.net

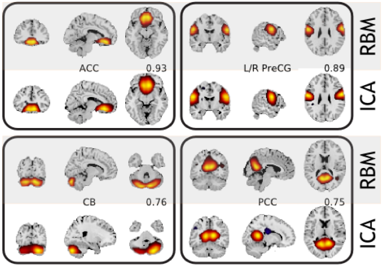
Summary (the claims of deep learning hold)

- probabilistic building block works competitively[a]
- classification performance is great! and holds promise
- the deep belief net learns more with depth
- the overall approach has high potential for facilitating discovery in our field

[a] Hjelm, R. D. *et al.* *NeuroImage* **96,** 245–260 (2014).

Summary (the claims of deep learning hold)

- probabilistic building block works competitively[a]
- classification performance is great! and holds promise
- the deep belief net learns more with depth
- the overall approach has high potential for facilitating discovery in our field



[a] Hjelm, R. D. *et al.* *NeuroImage* **96,** 245–260 (2014).

Summary (the claims of deep learning hold)

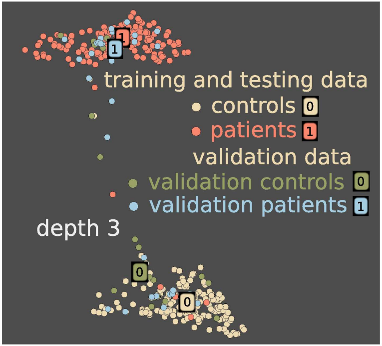
- probabilistic building block works competitively[a]
- classification performance is great! and holds promise
- the deep belief net learns more with depth
- the overall approach has high potential for facilitating discovery in our field

depth	3
SVM F-score	0.90 ± 0.14
LR F-score	0.91 ± 0.14
KNN F-score	0.90 ± 0.16

[a] Hjelm, R. D. *et al.* *NeuroImage* **96,** 245–260 (2014).

Summary (the claims of deep learning hold)

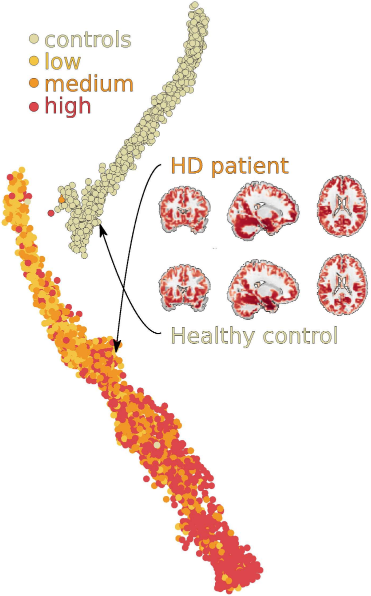
- probabilistic building block works competitively[a]
- classification performance is great! and holds promise
- the deep belief net learns more with depth
- the overall approach has high potential for facilitating discovery in our field



[a] Hjelm, R. D. *et al.* *NeuroImage* **96,** 245–260 (2014).

Summary (the claims of deep learning hold)

- probabilistic building block works competitively[a]
- classification performance is great! and holds promise
- the deep belief net learns more with depth
- the overall approach has high potential for facilitating discovery in our field



[a] Hjelm, R. D. *et al.* *NeuroImage* **96,** 245–260 (2014).

Future work

- Learn to learn dynamic repesentations (temporal models)
- Take advantage of already developed multimodal approaches
- Better understand parameterization for various types of data
- Extend deep models to better model independence and nonnegativity
- Apply deep learning to unsolved or difficult problems
- Many more...

Thank you!